

APPLICATION FOR UNITED STATES LETTERS PATENT

INVENTORS: Hung-Jin KIM

TITLE: NON-VOLATILE SEMICONDUCTOR MEMORY DEVICE AND  
FABRICATING METHOD THEREOF

ATTORNEYS: FLESHNER & KIM, LLP  
& P. O. Box 221200  
ADDRESS: Chantilly, VA 20153-1200

DOCKET NO.: HME/D-010

# NON-VOLATILE SEMICONDUCTOR MEMORY DEVICE AND FABRICATING METHOD THEREOF

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

[1] The present invention relates to a non-volatile semiconductor memory device and a fabricating method thereof, more particularly, to a flash/EEPROM in a semiconductor device and a fabricating method thereof.

### 2. Background of the Related Art

[2] EEPROM is characterized in two categories: a byte erasure type and a flash type. The flash type is further divided into a hot-electron injection type and a F-N(Fowler-Nordheim) current type. An EEPROM/flash type memory device is basically constructed with a MOS transistor having stacked polysilicon gates including a floating gate and a control gate. When there are no electrons in the floating gate, a channel for current to flow through is generated as voltage is applied to the control gate and the voltage applied to the control gate exceeds a threshold voltage to form the channel. When electrons are accumulated in the floating gate, the voltage applied to the control gate forms the channel at a higher level than when there are no electrons in the floating gate because of an electric field offset effect due to the negative charges in the floating gate.

[3] A semiconductor non-volatile memory operates by the principle that electrons accumulated in the floating gate will be unable to escape when an energy barrier exists in both the silicon substrate and the control gate. Electron injection and discharge are performed on a floating gate of a semiconductor non-volatile memory to make the memory electrically rewritable where the threshold voltage of a memory cell increases proportionally with the amount of injected electrons. A non-volatile memory in a semiconductor device is enabled by designating an electron-accumulated state as '1' and a non-electron state as '0' because electrons accumulated around the floating gate do not escape and external electrons do not enter into the floating gate.

[4] A basic structure of a stacked gate, which introduces no selection gate, is similar to a UV-erasing type memory device and is particularly useful for fabricating a highly-integrated memory device. Erasing is achieved by having a control gate, which is either grounded or receiving a negative voltage, and applying a high voltage to a source, resulting in a F-N current between the source and a floating gate thus decreasing the threshold voltage. Alternatively, a negative voltage may be applied to the substrate for a blanket erasing.

[5] Source/drain regions of a non-volatile memory device such as an EEPROM are impurity diffusion regions doped heavily with impurities, while a channel region is formed in an active area of a substrate over which the floating and control gates overlap each other.

[6] Cell programming is achieved by injecting hot channel electrons, which are generated from a drain stage of the channel region by applying a predetermined voltage to the control gate and the drain region, respectively, into a floating gate. Programming is achieved by hot channel electron injection so that the hot carriers generated near the drain are injected into the floating gate from the drain, provided that the channel region is inversed by applying a high voltage to the control gate and applying a proper voltage to the drain, where the source and well are grounded. Erasing is achieved by discharging carriers stored in the floating gate. Erasing of an EEPROM is completed by U-V erasure or by applying a high voltage to the source or drain regions or the bulk. Reading is achieved by judging a cell status of on/off by means of reading the threshold voltage of a cell transistor. For instance, once the control gate and the drain are supplied with 5V and 1V respectively, the threshold voltage of the programmed cell becomes high (at least over 5V) to be 'off', while the erased cell becomes low to be 'on'. In other words, EEPROM (electrically erasable and programmable read only memory) enables data to be programmed and erased electrically, which means that EEPROM installed in a system enables the system to rewrite data with ease.

[7] Flash EEPROM developed from EEPROM enables programming of data by a byte or more according to the design and to erase data by bits or a block, thereby improving the operating speed of EEPROM greatly. Accordingly, a non-volatile flash memory device, of which integration is superior to that of EEPROM, meets the needs of

a high capacity memory device and enables high speed data reading operation for portable office automation. However, a flash memory device of 1 Tr type (ETOX) fabricated by the related art consumes electric current excessively, thereby requiring an oversized voltage pump circuit. And, the flash memory device according to the related art requires an additional circuit for preventing electrical disturbance if a selection gate is not introduced. Moreover, the more frequently the device is used, the less reliable the flash memory device according to the related art becomes because of the limited endurance of programming and erasing.

[8] Figure 1A and Figure 1B show cross-sectional views seen from the directions of channel length and width, respectively, of an EEPROM device in a semiconductor device fabricated by a related art. In Figures 1A-1B, a field oxide layer 11 defining a device isolation area and a device active area are formed by LOCOS (local oxidation of silicon) on predetermined portions of a silicon substrate 10, thereby exposing the device active area. In the active area of the substrate 10, a gate insulating layer 12 formed of oxide, a floating gate 13 formed of polysilicon on the gate insulating layer 12, an inter-poly layer 14 formed of an insulator of an O-N-O structure, and a control gate 15 of polysilicon are formed.

[9] A method of fabricating a non-volatile memory device according to the related art includes forming a second conductive type well in a predetermined portion of a first conductive type silicon substrate 10, then forming by LOCOS a field oxide layer

11, which acts as a device isolation layer. Next, a gate oxide layer 12 is formed by oxidizing a surface of the substrate 10 thermally and a polysilicon layer is deposited on the gate oxide layer 12, followed by the formations of a floating gate 13 by patterning the polysilicon layer and the gate insulating layer so that the polysilicon layer and the gate insulating layer remain within a memory cell only. An O-N-O layer 14 is then formed on an exposed top surface of the floating gate 13.

[10] After an upper polysilicon layer has been deposited on the substrate including the surface of the O-N-O layer 14, where the O-N-O layer 14 acts as an insulating layer between the polysilicon layers, a control gate 15 is formed in a direction along the length of the channel by patterning the upper polysilicon layer to extend to another memory cell. Then, an impurity diffusion region 16 is formed by implanting first conductive type impurity ions, where the control gate 15 is used as a mask, then diffusing the impurity ions to form source and drain regions 16. In this case, the impurity diffusion regions 16 may be formed with either a symmetrical junction or an asymmetrical junction, where the symmetrical junction is used for a negative erase while the asymmetrical junction used for a positive erase.

[11] As mentioned in the above description of the non-volatile memory device (of a single poly gate type EEPROM) fabricated by the related art, the area occupied by a cell with a single gate is relatively large and the voltage applied for programming and erasing operations carried out by F-N tunneling is relatively high. Moreover, in a split

gate type EEPROM constructed with 2 or 3 gates according to the related art, programming and erasing are achieved by the injection of hot electrons and by F-N tunneling, respectively. Namely, the split gate type introduces a structure where the control gate covers the floating gate.

[12] As described above, the related art EEPROM have various disadvantages. The split-gate type EEPROM according to the related art produces disturbances and consumes lots of power when a programming is carried out by hot electron injection. Additionally, in a non-volatile memory cell according to the related art, it is difficult to increase a coupling ratio, which is defined as the areas of the control gate and the floating gate in contact with the inter-poly layer 14 where the coupling ratio depends on a planar structure that does not include the part over the field oxide layer 11.

### SUMMARY OF THE INVENTION

[13] An object of the present invention is directed to a non-volatile semiconductor memory device and a fabricating method thereof that substantially obviate one or more of the problems due to limitations and disadvantages of the related art.

[14] Another object of the present invention is to provide a flash/EEPROM in a semiconductor device and a fabricating method thereof that prevents programming disturbance and enables programming operated by a byte unit by achieving a

programming and an erasing of a memory device through a F-N tunneling using a coupling ratio between floating and control gates.

[15] Another object of the present invention is to provide a flash/EEPROM in a semiconductor device and a fabricating method thereof that has a floating gate over a control gate having a source region less overlapped than a drain region.

[16] Another object of the present invention is to provide a flash/EEPROM in a semiconductor device and a fabricating method thereof that reduces a cell area by removing a control transistor.

[17] Another object of the present invention is to provide a flash/EEPROM in a semiconductor device and a fabricating method thereof that prevents programming disturbance and enables programming operated by a byte unit by achieving a programming and an erasing of a memory device through a F-N tunneling using a coupling ratio between floating and control gates by means of forming a structure where the control gate is covered with the floating gate, a drain region is properly overlapped by coupling, and a source region is less overlapped.

[18] Another object of the present invention is to provide a memory device compatible with a flash EEPROM by realizing various functions such as MCU on one chip.



[19] Another object of the present invention is to realize an ultra-highly integrated device by reducing a cell area itself and a total memory area by means of removing a control transistor for erasing by byte unit in an EEPROM.

[20] To achieve at least these and other advantages in a whole or in part and in accordance with the purpose of the present invention, as embodied and broadly described, the present invention includes a control gate on a semiconductor substrate in which a device active area and a device isolation are defined, the control gate lying on the device active area, the control gate dividing the device active area into a first region and a second region wherein a control gate insulating layer is inserted between the substrate and the control gate, a first insulating layer covering a top surface and a side of the control gate, a drain junction in the first region of the substrate, a source junction in the second region, the source junction separated from the control gate to a predetermined distance, a second insulating layer on the second region between the source junction and the control gate and on a predetermined portion of a surface of the first region between the first insulating layer and the drain junction, and a floating gate covering the second insulating layer and the first insulating layer.

[21] To further achieve the above objects in a whole or in part, a non-volatile semiconductor memory device according to the present invention can further include a third insulating layer at a later surface of the first insulating layer, an insulating interlayer covering the device active area including a surface of the floating gate, a contact hole in

a predetermined portion of the insulating interlayer, the contact hole exposing a portion of the drain junction in the first region where the floating gate is not located, and a bit line on the insulating interlayer, the bit line contacted with the drain junction through the contact hole.

[22] To further achieve the above objects in a whole or in part, a semiconductor device according to the present invention includes a plurality of device active areas separated from one another in parallel by a plurality of device isolation layers in a first direction on a semiconductor substrate where a plurality of memory cell areas are defined, a plurality of control gates crossing over the device active areas and the device isolation layers, where the respective control gates are separated from one another in a second direction perpendicular to the first direction, a plurality of drain junctions in the device active areas at one side of the respective control gates, respectively, a plurality of source junctions in the device active areas at the other side of the respective control gates, respectively, the respective source junctions are separated from the control gates by a predetermined distance, respectively, and a plurality of floating gates overlap with the control gates, portions of the drain junctions, and portions of the device active areas where the source junctions are not formed at the other sides of the control gates, respectively.

[23] To further achieve the above objects in a whole or in part, the lay out of the semiconductor device according to the present invention can further include a plurality

of bit line contacts at predetermined portions of the drain junctions which are not overlapped with the floating gates in the memory cell areas, respectively, and a plurality of bit lines contacted with the respective bit line contacts, the bit lines overlapped with the respective device active areas, the bit lines extending to the first direction.

[24] To further achieve the above objects in a whole or in part, a method of forming a semiconductor device according to the present invention includes forming a control gate on a predetermined portion of a memory cell area of a semiconductor substrate wherein the memory cell area which is divided into a first region and a second region by the control gate is defined by a device isolation layer and wherein a control gate insulating layer is inserted between the control gate and the semiconductor substrate, forming a first insulating layer on an exposed surface of the control gate, forming a drain junction in the first region, forming a third insulating layer covering predetermined portions of the second region and the drain junction wherein the third insulating layer is extended from the first insulating layer, forming a floating gate on the third insulating layer and an exposed surface of the first insulating layer, and forming a source junction in the second region which is not overlapped with the floating gate.

[25] To further achieve the above objects in a whole or in part, after the step of forming the first insulating layer, the method can further include forming a second insulating layer on a surface of the first insulating layer at a side of the control gate

wherein the second insulating layer differs from the first insulating layer greatly in etch selectivity.

[26] To further achieve the above objects in a whole or in part, the forming a third insulating layer and a floating gate can include forming an oxide layer on the exposed surface of the first region and a tunneling oxide layer on the second region of the substrate, forming a conductive layer over the substrate including the oxide layer, the tunneling oxide layer, and the first insulating layer, and patterning the conductive layer, the oxide layer and the tunneling oxide layer to be overlapped with the control gate and portions of the drain and source junctions.

[27] To further achieve the above objects in a whole or in part, a method of forming a semiconductor device according to the present invention includes the steps of forming a control gate on a predetermined portion of a memory cell area of a semiconductor substrate wherein the memory cell area which is divided into a first region and a second region by the control gate is defined by a device isolation layer and wherein a control gate insulating layer is inserted between the control gate and the semiconductor substrate, forming a first insulating layer on an exposed surface of the control gate, forming a second insulating layer on a surface of the first insulating layer at a side of the control gate wherein the second insulating layer differs from the first insulating layer greatly in etch selectivity, forming a drain junction in the first region, forming an oxide layer on the exposed surface of the first region and a tunneling oxide layer on the second

region of the substrate, forming a conductive layer over the substrate including the oxide layer, the tunneling oxide layer, and the first insulating layer, forming a floating gate and a third insulating layer by patterning the conductive layer, the oxide layer and the tunneling oxide layer to be overlapped with the control gate and portions of the drain and source junctions, and forming a source junction in the second region which is not overlapped with the floating gate.

[28] To further achieve the above objects in a whole or in part, after the forming the source junction in the second region, the method can further include forming an insulating layer over the semiconductor substrate including the floating gate, forming a contact hole exposing a portion of the drain junction which is not covered with the floating gate by removing a portion of the insulating layer, and forming a bit line on the insulating interlayer wherein the bit line fills up the contact hole.

[29] To further achieve the above objects in a whole or in part, in a non-volatile semiconductor memory device including a plurality of memory cells each of which consists of a control gate, a drain junction, a source junction, a floating gate, a word line, and a bit line, wherein an electric charge  $Q1$  of a first parasitic capacitor generated from a voltage  $V_{cg}$  applied to the control gate and a voltage  $V_{fg}$  induced on the floating gate is  $C1(V_{cg} - V_{fg})$  and  $C1$  is a capacitance of the first parasitic capacitor, wherein an electric charge  $Q3$  of a third parasitic capacitor generated from a voltage  $V_s$  at the source junction and a voltage  $V_{fg}$  induced on the floating gate is  $C3(V_s - V_{fg})$  and  $C3$  is a capacitance of

the third parasitic capacitor, wherein an electric charge  $Q_4$  of a fourth parasitic capacitor generated from a voltage  $V_b$  at a bulk and a voltage  $V_{fg}$  induced on the floating gate is  $C_4(V_b - V_{fg})$  and  $C_4$  is a capacitance of the fourth parasitic capacitor, and wherein an electric charge  $Q_2$  of a second parasitic capacitor generated from a voltage  $V_d$  at the drain junction and a voltage  $V_{fg}$  induced on the floating gate is  $C_2(V_d - V_{fg})$  and  $C_2$  is a capacitance of the second parasitic capacitor, to further achieve the above objects in a whole or in part, a method of preventing disturbance of a non-volatile semiconductor memory according to the present invention includes such that (1) a programming is achieved by selecting one of two variables  $V_{cg}$  and  $V_d$  which prevents the disturbance at the word line and the bit line according to the formula  $V_{fg} = (C_1 V_{cg} + C_2 V_d) / C_{total}$  and wherein the programming is achieved by F-N tunneling using voltage difference between the floating gate and the source junction and a channel region at a side of the source junction, thereby injecting electrons into the floating gate, and (2) an erasing is achieved by discharging electrons accumulated in the floating gate in use of F-N tunneling, which uses voltage difference between the floating gate and a lateral side of the source junction, and wherein the erasing is achieved by selecting the voltages  $V_{cg}$  and  $V_s$  not to discharge the electrons into the source junction due to a voltage applied to another memory cells sharing the word line. Preferably,  $V_{fg}$  is equal to  $[C_1 V_{cg} + C_2 V_d + C_3 V_s + C_4 V_b] / C_{total}$  as total of the electric charges ( $Q_1 + Q_2 + Q_3 + Q_4$ ) of the first to fourth parasitic

capacitors is approximately equal to 0 when the floating gate at neutral and  $C_{total}$  is approximately equal to  $C_1 + C_2 + C_3 + C_4$ .

### BRIEF DESCRIPTION OF THE DRAWINGS

[30] The invention will be described in detail with reference to the following drawings in which like reference numerals refer to like elements wherein:

[31] Figs. 1A and Figs. 1B show cross-sectional views seen from the directions of channel length and width respectively, of a flash EEPROM device in a semiconductor device fabricated by a related art;

[32] Figs. 2A and Figs. 2B are diagrams that show a cross-sectional view seen from the direction of channel length and a layout of an EEPROM device in semiconductor devices fabricated by preferred embodiments of the present invention;

[33] Fig. 3 is a diagram that shows a circuit of a non-volatile memory device according to a preferred embodiment of the present invention;

[34] Fig. 4 is a diagram that shows a circuit for capacitors in a non-volatile memory device according to a preferred embodiment of the present invention; and

[35] Figs. 5A to Figs. 5D are diagrams that show cross-sectional views from a channel length direction of fabricating a semiconductor memory device according to a preferred embodiment of the present invention.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[36] Preferred embodiments of the present invention relate to a flash/EEPROM in a semiconductor device and a fabricating method thereof that reduce or prevent programming disturbances and enable programming and erasing of a memory device through F-N tunneling using a coupling ratio between floating and control gates, which enables programming operated by a byte unit to be applied to an EEPROM or a flash EEPROM, where the coupling ratio between a floating gate capacitor and a control gate capacitor determines a control gate voltage applied to a floating gate as well as the programming time.

[37] In one case, the coupling ratio that determines a voltage level in the floating gate has an influence on a threshold voltage  $V_{Tpgm}$  level of the programming. Compared to a single poly gate type structure of the related art, the present invention can preferably add a control gate to improve the coupling ratio and reduce an area occupied by a memory cell by means of negative erasure. Further, compared to a split gate type non-volatile memory device of the related art, the present invention can use a structure where a floating gate lies over a control gate. Thus, according to the present invention, programming can be executed by F-N tunneling and any disturbances are prevented without needing an additional inhibition-preventing mode.

[38] Fig. 2A is a diagram that shows a cross-sectional view seen from a first direction, which is the direction of channel length of a non-volatile semiconductor



memory device fabricated by the present invention, where a gate insulating layer 610 of silicon oxide or the like and a control gate 62 are formed in order on a top surface of a device active area of a silicon substrate 60 of a first conductive type in which device active and isolation areas are defined, respectively. In a preferred embodiment, the control gate 62 is elongated in a second direction, which is a channel width direction of the device, and is preferably coupled to other control gates of neighboring memory cells (not shown). The control gate 62 is formed of polysilicon doped with first or second conductive type impurity ions for electric conduction.

[39] An oxide layer 64 for an inter-polysilicon layer 64 is preferably formed to a predetermined thickness to electrically isolate the control gate 62 from a floating gate 69 on a top surface and a side of the control gate 62 where the gate insulation layer 610 has not been formed. A sidewall spacer 65 of a nitride layer or the like is preferably formed at a lateral side of the inter-polysilicon layer 64 of oxide. Next, a drain 67 of a second conductive type impurity ion diffusion region or the like is formed in a drain forming region of the substrate 60, which lies at and extends from one end of the control gate 62. As shown in Figure 2A, centering around the control gate 62, a source 72 of a second conductive type impurity ion diffusion region is formed in a source forming region of the substrate separated from the control gate 62 by a predetermined distance and confronts the drain forming region in the device active area of the substrate 60.

[40] A floating gate 69 is preferably formed on surfaces of the sidewall spacer 65 of nitride and the inter-polysilicon layer 64. One end of the floating gate 69 extends over a portion of the drain 67, while the other end of the floating gate 69 extends to an edge of the source 72. In this case, a tunneling oxide layer 68 is inserted between the floating gate 69 and the device active area. Separated from other floating gates in the neighboring memory cells, the floating gate 69 is designed to remain inside a single memory cell and is formed of a polysilicon doped with a first or second conductive type impurity for electric conduction.

[41] An oxide layer 71 is formed on a surface of the floating gate 69 and the remaining surfaces of the active area preferably by oxidation to cover the exposed surfaces of the drain 67 and the source 72, as well as, the exposed surface of the floating gate 69. A transistor for a non-volatile memory device is preferably constructed with the control gate 62 surrounded by the insulating layers 610, 64, and 65, the tunneling oxide layer 68, the floating gate 69, and the source 72 and the drain 67. An insulating interlayer 73 is then formed of oxide or the like over the substrate, including the transistor, to be used as a non-volatile memory device. A bit line 74 is finally formed on a surface of the insulating interlayer 73 and electrically coupled to a portion of the drain 67, thereby completing a semiconductor non-volatile memory device.

[42] Fig. 2B shows a layout of an EEPROM device, which is similar to Figure 2A, in a semiconductor device where a device isolation area 700 and device active areas are

defined on a silicon substrate of a first conductive type semiconductor substrate by a field oxide layer 700 elongated to a first direction, which is horizontal in Figure 2B. A control gate 62, beneath which a control gate insulating layer of oxide lies, is formed on a surface of the active area of the silicon substrate. A bit line 74, which penetrates into a portion of an insulating interlayer 73 and is then electrically coupled through a bit line contact 740 to a portion of a drain 67 that is not overlapped with a floating gate 69 or an oxide layer 71, is formed on a surface of the insulating interlayer 73 (not shown), thereby completing a semiconductor non-volatile memory device.

[43] The device active areas are defined by a plurality of the field oxide layers 700, running in a first direction and a plurality of the control gates 62 are elongated in a second direction, or the channel width direction, crossing the device active areas and the plurality of field oxide layers 700.

[44] Each memory cell shares the source 72 and the drain 67 of the adjacent left and right memory cells in the first direction, respectively. The floating gate 69 covers over the control gate 62, preferably has a rectangular shape and is formed in a domain of the single memory cell area. As illustrated in Figure 2B, for example, the drain 67 partially overlaps the floating gate 69, while the source 72 is less overlapped with the floating gate 69.

[45] Fig. 3 is a diagram that shows a schematic circuit of a single cell according to a preferred embodiment of the present invention of a non-volatile memory device or

the like where the voltages applied to a control gate, a drain, a source, and a substrate bulk are designated by  $V_{cg}$ ,  $V_d$ ,  $V_s$ , and  $V_b$ . A floating gate can be coupled between the control gate and the substrate bulk, to allow data to be stored electrically by storing or discharging electrons. Programming or erasing can be controlled by applying a predetermined voltage to the control gate where the control gate functions as a gate of a transistor on reading. On programming or erasing, electrons can be injected into or discharged from the floating gate from a lateral side of the source and the channel regions, while a tunneling oxide layer, through which F-N current flows, is inserted between the source and the floating gate.

[46] Fig. 4 is a diagram that shows a circuit for capacitors in a non-volatile memory device to preferably illustrate parasitic capacitances in a preferred embodiment where an electric charge  $Q_1$  of a parasitic capacitor, which is generated between a control gate with an applied voltage of  $V_{cg}$  and a floating gate with an induced voltage of  $V_{fg}$  becomes  $C_1(V_{cg} - V_{fg})$ , where  $C_1$  is a capacitance of the parasitic capacitor. Another electric charge  $Q_3$  of a parasitic capacitor generated between a floating gate with a voltage of  $V_{fg}$  and a source with a voltage of  $V_s$  becomes  $C_3(V_s - V_{fg})$ , where  $C_3$  is a capacitance of the corresponding parasitic capacitor.

[47] Another electric charge  $Q_4$  of a parasitic capacitor generated between a floating gate with a voltage of  $V_{fg}$  and a bulk substrate with a voltage of  $V_b$  becomes  $C_4(V_b - V_{fg})$ , where  $C_4$  is a capacitance of the corresponding parasitic capacitor. A further

electric charge  $Q_2$  of a parasitic capacitor generated between a floating gate with a voltage of  $V_{fg}$  and a drain with a voltage of  $V_d$  becomes  $C_2(V_d - V_{fg})$ , where  $C_2$  is a capacitance of the corresponding parasitic capacitor. From such relations, when the floating gate is neutral, the total electric charges ( $Q_1 + Q_2 + Q_3 + Q_4$ ) is 0. In other words,  $V_{fg} = [C_1V_{cg} + C_2V_d + C_3V_s + C_4V_b]/C_{total}$ , wherein  $C_{total}$  is  $C_1 + C_2 + C_3 + C_4$ . When the voltages applied to the source and bulk are 0 (that is,  $V_s = V_b = 0$ ), the voltage of the floating gate is denoted by  $V_{fg} = (C_1V_{cg} + C_2V_d)/C_{total}$  or  $V_{fg} = (C_1V_{cg} + C_3V_s)/C_{total}$ .

[48] A coupling ratio of an EEPROM, as explained in the previous description, is numerally expressed by the degree of a voltage induced on the floating gate when a bias voltage is applied to a single cell. Therefore, preferred embodiments according to the present invention reduce or prevent neighboring cells from malfunctions such as data loss, disturbance on programming or erasing by using the coupling ratio.

[49] Operations of an EEPROM, according to preferred embodiments of Figures 3 and 4 will now be described, where initially for programming an EEPROM,  $[V_{cg} = V_{pp1}, V_d = V_{pp2}, V_b = 0, V_s = V_{fg}]$ ,  $[V_{cg} = V_{pp1}, V_d = 0, V_b = 0, V_s = V_{fg}]$ , and  $[V_{cg} = 0, V_d = V_{pp2}, V_b = 0, V_s = V_{fg}]$  are programmed into a selected cell, a word line sharing cell, and a bit line sharing cell, respectively. Programming uses the phenomenon that occurs when electrons are injected into the floating gate by F-N tunneling effect, which uses a voltage difference between the floating gate and the source and channel region at the side of the source. In this case,  $V_{fg}$  is determined by the formula of

$V_{fg} = (C_1 V_{cg} + C_2 V_d) / C_{total}$ , which depends on a couple of variables  $V_{cg}$  and  $V_d$ , wherein the disturbance generated from the word line sharing cell is a factor of  $V_{fg} = C_1 V_{cg} / C_{total}$ , which is lower than another  $V_{fg}$  of a cell to be programmed. Therefore, a value of  $V_{cg}$  for preventing the disturbance may be selected, and other disturbances by the bit line may be prevented by similar methods.

[50] A preferred embodiment according to the present invention can prevent such disturbances and thus meets the requirement for a byte erasable EEPROM sufficiently. As electrons are injected into a floating gate of a programming cell, the threshold voltage of a transistor is increased by the floating gate on the source region of the cell. Namely, on reading,  $V_{fg} < V_T$  is satisfied by  $V_{cg}$  and  $V_d$ . Initially for erasing an EEPROM,  $[V_{cg} = (-V_{pp1}), V_s = V_{pp3}, V_d = 0, V_b = 0]$ ,  $[V_{cg} = (-V_{pp}), V_d = 0, V_b = 0, V_s = 0]$ , and  $[V_{cg} = 0, V_d = 0, V_b = 0, V_s = V_{pp}]$  are programmed into a selected cell, a word line sharing cell, and a source sharing cell, respectively.

[51] Erasing is achieved by the negative erasing method, wherein electrons accumulated in the floating gate are discharged by F-N tunneling, which uses the voltage difference between the floating gate and the junction region at the side of the source. In this case, as the word line sharing cells receive stresses amounting to  $-V_{pp}$ , which is a prescribed voltage of  $-V_{pp}$  that prevents electrons from being discharged into the source region by the voltage of  $-V_{pp}$ , and thus prevents a disturbance from being generated if  $V_{pp3}$  is applied to the source line. Thus, the threshold voltage of the transistor including

the floating gate near the side of source is decreased under negative value by the discharged electrons, thereby becoming a sufficient inversion state.

[52] Figures 5A to 5D illustrate a preferred method of the present invention, which can, for example, be used to form the preferred embodiment shown in Figure 2. As shown in Figure 5A, a device isolation layer is formed on a silicon substrate 60 of a first conductive type to define a device isolation area and a device active area by STI (shallow trench isolation), LOCOS (local oxidation of silicon) or the like. Next, the threshold voltage of a device is adjusted by ion implantation on an exposed surface of the substrate 60. An oxide layer 61 for a gate insulating layer is formed by thermal oxidation or the like on a surface of the active area of the exposed silicon substrate 60. Then, a conductive layer for a control gate is formed on the oxide layer 61 by chemical vapor deposition or the like. In this case, the conductive layer is formed by depositing in-situ doped polysilicon or by depositing undoped polysilicon and then doping the undoped polysilicon with impurities for electric conduction.

[53] Compared to the steps of forming a floating gate and forming a control gate beneath which an inter polysilicon layer is formed according to the related art, a preferred method of the present invention forms an etch mask 63 on the conductive layer for a control gate. In this case, the etch mask 63 forms a control gate by coating the conductive layer for a control gate with photoresist and then by defining a photoresist pattern 63 by exposure and development of the photoresist, where the etch mask preferably extends to

other portions of the conductive layer in the neighboring cells. Then, a control gate 62 is formed of doped polysilicon by removing a portion of the conductive layer not covered with the etch mask 63 by anisotropic etching such as dry etching and the like, thereby exposing a portion of the oxide layer 61 for forming a gate insulating layer.

[54] In this case, one of the control gate 62 extends to other control gates of neighboring cells over the device active and isolation areas in a predetermined direction, and one side of the device active area is a source forming region while the other side is a drain forming region centering around the control gate 62. As shown in Fig. 5B, a control gate insulating layer 610 including a remaining oxide layer 610 is formed by removing a portion of the oxide layer that is not covered with the etch mask using the anisotropic etch such as dry etch or the like. Thus, the control gate insulating layer 610 is formed between the control gate 62 and the substrate 60, and the etch mask formed of the remaining photoresist is removed by  $O_2$  action or the like, thereby exposing a surface of the control gate 62. Next, an insulating layer 64 is formed on the exposed surface of the control gate 62. In this case, the insulating layer 64 is formed preferably by oxidizing the exposed surface of the control gate 62 of polysilicon preferably by thermal oxidation or the like. Then, a nitride layer is formed over the substrate, including the insulating layer of oxide, and a sidewall spacer 65 is formed by leaving a portion of the nitride layer 65 at the lateral sides of the insulating layer 64 by using a surface of the oxide layer on the top surface of the control gate 62 as an etch-stop layer.



[55] Thus, the control gate insulating layer 610 separates the control gate 62 from the substrate 60, the insulating layer 64 of oxide covers the remaining surface of the control gate 62 for isolation, and the nitride of the sidewall spacer 65 insulates the sidewall spacer 65 on at least one of the lateral sides of the control gate 62 and protects at least one of the lateral side surfaces of the control gate 62, thereby preferably protecting the insulating layer 64 and the control gate insulating layer 610 from damage by the cleaning process before forming a tunneling oxide layer.

[56] Next, the substrate, including the control gate protected by the insulating layer 64, is coated with photoresist and the photoresist is patterned by exposure and development. The photoresist pattern 66 preferably exposes the drain forming region of the device active area but may also be patterned to expose other regions. In this case, the drain and source forming regions of the device active area preferably oppose each other and are centered around the control gate 62, thereby improving a process margin for forming the photoresist pattern 66 for exposing the drain forming region. Then, a second conductive type impurity ion buried layer is formed by implanting second conductive type impurity ions I1 heavily into the drain forming region of the exposed substrate, and a drain 67 is formed by diffusion of the second conductive type impurity ion buried layer preferably by heat treatment such as annealing or the like. In this case, the photoresist pattern 66 may be removed before the diffusion or by  $O_2$  action after the formation of the drain 67.

[57] As shown in Fig. 5C, the photoresist pattern is removed and a cleaning process is carried out on the entire surface of the substrate including the above structure, which leaves the source forming region and a surface of the drain 67 exposed. Then, an oxide layer 68 for a tunneling oxide layer 68 is formed on the exposed surface of the device active area of the silicon substrate 60. In this case, the tunneling oxide 68 may be formed by oxidizing the surface of the substrate of silicon thermally, wherein a portion of the oxide layer 68 on the source forming region is formed thinner and preferably formed to be 4 times thinner than the other portion of the oxide layer 68 formed on the drain forming region. Next, another conductive layer 69 is formed on the insulating layer 64, the sidewall spacer 65, and the tunneling oxide layer 68 to form a floating gate 69. In this case, the conductive layer 69 to form a floating gate is preferably formed by depositing in-situ doped polysilicon or by depositing undoped polysilicon and then doping the undoped polysilicon with impurities for electric conduction.

[58] After the conductive layer 69 for a floating gate has been coated with photoresist, a photoresist pattern 70 is formed by exposure and development for use as an exposure mask for forming the floating gate. In this case, the photoresist pattern 70 is formed to remain partially on the conductive layer 69 over the control gate 62, a portion of the drain 67, and a portion of the source forming region. Then, the floating gate 69 and the tunneling oxide layer 68 are formed by sequentially removing the portions of the conductive layer 69 not needed and the portions of the oxide layer not needed, preferably

by anisotropic etch such as dry etch or the like. After removal, the floating gate 69, which is the remaining conductive layer, can overlap the control gate 62 and can extend partially into the drain 67 and the source forming regions which in turn can cause the floating gate 69 to remain within the cell area and separated from the neighboring cells.

[59] After the photoresist pattern for forming the floating gate has been removed preferably by  $O_2$  action or the like to expose a surface of the floating gate 69, as illustrated in Figure 5D, the surface of the floating gate 69 preferably undergoes a cleaning process to remove particles. An insulating layer 71 can then be formed on the exposed surface of the floating gate 69. In a preferred embodiment, the insulating layer 71 is preferably formed by oxidizing the exposed surface of the floating gate 69 and the exposed surfaces of the drain 67 and the source forming region. Next, a source 72 is preferably formed by doping the exposed source forming region of the substrate with second conductive type impurity ions, where the source 72 is a second conductive type doping region.

[60] The source 72 is preferably formed by forming a photoresist pattern 73 exposing the source forming region after the oxide layer 71 has been coated with photoresist. In this case, a process margin of forming the photoresist pattern 73 exposing the source forming region is sufficient since the drain forming region opposes the source forming region in the device active area centering around the control gate 62 on opposite sides. A second conductive type impurity ion buried layer can be formed by implanting second conductive type impurity ions  $I_2$  heavily into the source forming region of the

exposed substrate. Then, the source 72 can be formed by diffusing the second conductive type impurity ion buried layer preferably by heat treatment such as annealing or the like. In this case, the photoresist pattern 73 may be removed before the diffusion or by  $O_2$  action after the formation of the source 72.

[61] Thus, the control gate 62 can overlap the drain 67 in part and is preferably separated from and does not overlap the source 72. The floating gate 69 preferably overlaps the tunneling oxide layer 68 and preferably overlaps a large portion of the drain 67 and a small portion of the source 72. Thereafter, a surface of the oxide layer 71 can then be exposed again by removing the photoresist pattern 73 for forming the source 72. After an insulating interlayer may be formed on the substrate including the above structure and a contact hole exposing a portion of the drain 67, which is not overlapped with the floating gate 71, may also be formed by removing a portion of the insulating interlayer and the oxide layer 71. Then, a plug may be formed by filling the contact hole and a bit line may be coupled to the plug in or on the insulating interlayer, thereby completing a fabrication of a non-volatile memory cell in a semiconductor device such as EEPROM, flash/EEPROM and the like.

[62] As described above, preferred embodiments of a memory cell EEPROM and methods for forming same have various advantages. Preferred embodiments according to the present invention provide a flash EEPROM able to erase data by block units or in its entirety and also provides a byte erasable EEPROM able to erase data by byte units.

The present invention also improves device integration by reducing an area occupied by each of the memory cells. Additionally, the preferred embodiments of the present invention consume a low amount of current since the programming and the erasing can be achieved by F-N tunneling, thereby reducing power consumption.

[63] Also, the preferred embodiments of the present invention are capable of consuming less electricity and therefore can be used with a single power source or low voltage operated products. Further, the preferred embodiments of the present invention are capable of charge pumping. Moreover, the preferred embodiments of the present invention can be programmed at a low level since the threshold voltage for programming is low enough not to form a channel on reading and the threshold voltage for erasing has sufficient negative value, thereby enabling the preferred embodiments of the present invention the capability of use with a single power source and low voltage operated products.

[64] It will be apparent to those skilled in the art that various modifications and variations can be made in a non-volatile semiconductor memory device and a fabricating method thereof of the present invention without departing from the spirit or scope of the inventions. Thus, it is intended that the present invention cover the modifications and variations of this invention provided they come within the scope of the appended claims and equivalents.